

UNCLASSIFIED

STANFORD UNIV CA DEPT OF STATISTICS

1/1

F J O'REILLY ET AL. 23 JUL 85 TR-361 N00014-76-C-0475

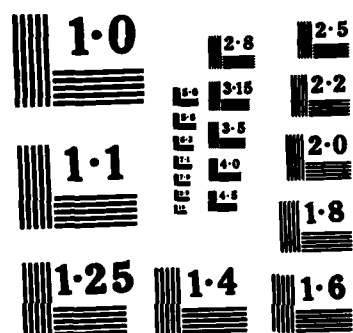
F/G 12/1

NL

END

FILMED

0740



NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

AD-A157 957

TRANSFORMING CENSORED SAMPLES AND TESTING FIT

BY

F. J. O'REILLY and M. A. STEPHENS

TECHNICAL REPORT NO. 361

JULY 23, 1985

Prepared Under Contract

N00014-76-C-0475 (NR-042-267)

For the Office of Naval Research

Herbert Solomon, Project Director

Reproduction in Whole or in Part is Permitted
for any purpose of the United States Government

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA



Accession For	
NTIS GRA&I <input checked="" type="checkbox"/>	
DTIC TAB	
Unannounced	
Justification	
By	
Distribution/	
Availability	
Dist	Availability Special
A-1	

1. INTRODUCTION.

In this article we consider tests of fit for a completely specified continuous distribution $F(x)$, where the x -sample is type I- or type II-censored. The censoring may be done in various ways: left- or right-censoring, as is often found in practice, also double-censoring, or censoring resulting from deletion of certain order statistics from the original sample.

The technique proposed generalizes a method earlier introduced by Michael and Schucany (1979) in the sense that it applies to an arbitrary censored sample and reproduces their results when the censoring is of type II and at one end only. The method introduced below is based on a transformation due to Rosenblatt (1952) and its inverse. All theoretical results are derived from the properties of these transforms, making it unnecessary to analyze each different kind of censoring separately.

Two procedures are obtained following the general methodology provided by the application of Rosenblatt's transformations. They transform the original censored data into a complete sample of ordered uniforms if the hypothesized distribution is correct. Therefore, the distributional test for the censored sample becomes a test that a full sample is uniform. There are many well-known methods of performing the latter test.

In Section 2, Rosenblatt's transformation is analyzed and its inverse introduced. In Section 3, the procedures are obtained for the type II censoring, and shown to be based on simple straight forward calculations. Some numerical examples are given. In Section 4, we examine for the type II censoring, how well the transformations work in terms of

power, when followed by Anderson-Darling's A^2 test for uniformity. The transformations discussed, followed by A^2 are found to give good results and are therefore recommended in view of their versatility.

In Section 5, an analysis is done for the type I censoring and the procedures are obtained.

Finally in Section 6, some general comments are given.

2. ROSENBLATT TRANSFORMS AND THEIR INVERSES.

It is well known that a test of fit of a completely specified continuous distribution is equivalent to a test for uniformity. If the original sample is censored, so that only a subset of the ordered sample is available, the equivalent problem is that of testing uniformity from the corresponding subset of the uniform ordered sample.

2.1. The Rosenblatt transformation.

If a set of random variables is given, consisting of independent variables whose marginal distributions are known and continuous, they can be transformed to a set of independent uniforms by mapping each with its corresponding distribution function. This is a direct application of the Probability Integral Transformation (PIT). If, however, the given variables are not independent, but their joint distribution is nevertheless known and absolutely continuous, then a transformation due to Rosenblatt (1952) can be applied that still yields independent uniforms. Specifically, let Y_1, Y_2, \dots, Y_m be jointly distributed with distribution function $G(Y_1, Y_2, \dots, Y_m)$, absolutely continuous.

Let $G_1(y_1)$ be the marginal distribution of Y_1

$G_{2/1}(y_2/y_1)$ be the conditional distribution of Y_2 given Y_1

.

$G_{m/1, \dots, m-1}(y_m/y_1, y_2, \dots, y_{m-1})$ be the conditional

distribution of Y_m given Y_1, Y_2, \dots, Y_{m-1}

then

$$U'_1 = G_1(Y_1)$$

$$U'_2 = G_{2/1}(Y_2/Y_1)$$

.

$$U'_m = G_{m/1, \dots, m-1}(Y_m/Y_1, \dots, Y_{m-1})$$

are independent variables, uniformly distributed in (0,1). Symbolically, this transformation is represented by:

$$\xrightarrow{Y_1, Y_2, \dots, Y_m} \boxed{R} \xrightarrow{U'_1, U'_2, \dots, U'_m} .$$

Observe that when applying R in general, one could select any of the $m!$ different orderings of the integers $1, 2, \dots, m$. We shall concentrate on the R -transformation applied to ordered uniform random variables, to produce new independent uniforms. In particular, only two orderings will be considered; the forward ordering given by the ordered uniforms and the backward ordering given by the ordered uniforms in reverse.

2.2. Application to ordered uniforms.

Let $U_{(1)}, U_{(2)}, \dots, U_{(n)}$ be a complete ordered sample of uniforms. Their joint distribution is of course known and absolutely continuous so that R can be applied. Denote this by R_F utilizing the forward ordering of the indices.

Symbolically,

$$\begin{array}{ccc} U_{(1)}, U_{(2)}, \dots, U_{(n)} & \xrightarrow{\boxed{R_F}} & U'_1, U'_2, \dots, U'_n \\ \text{ordered } U(0,1) \text{ sample} & & \text{unordered } U(0,1) \text{ sample} \end{array}$$

The variables U'_i can be shown to be

$$U'_i = 1 - \{(1 - U_{(i)}) / (1 - U_{(i-1)})\}^{n-i+1}, \quad i = 1, \dots, n \quad (2.1)$$

where $U_{(0)} \equiv 0$.

If the original sample of uniforms is ordered in reverse and R applied to it, we denote by R_B the resulting transformation applied to the backward-ordered uniform sample.

Symbolically

$$\begin{array}{ccc} U_{(1)}, U_{(2)}, \dots, U_{(n)} & \xrightarrow{\boxed{R_B}} & U''_1, U''_2, \dots, U''_n \\ \text{ordered } U(0,1) \text{ sample} & & \text{unordered } U(0,1) \text{ sample} \end{array}$$

The variables U''_i are given by

$$U''_i = (U_{(i)} / U_{(i+1)})^i, \quad i = 1, \dots, n \quad (2.2)$$

where $U_{(n+1)} \equiv 1$.

The transforms U' and U'' are well known, see for example Malmquist (1950).

2.3. Inverse transformations.

It is convenient to define R_F^{-1} and R_B^{-1} , the inverses of R_F and R_B . These transform a complete unordered uniform sample into a complete ordered uniform sample.

Let U_1', U_2', \dots, U_n' and $U_1'', U_2'', \dots, U_n''$ be two sets of independent $U(0,1)$ variables. Symbolically

$$\begin{array}{ccc} U_1', U_2', \dots, U_n' & \xrightarrow{\quad} & \boxed{R_F^{-1}} \xrightarrow{\quad} U_{(1)}, U_{(2)}, \dots, U_{(n)} \\ \text{unordered } U(0,1) \text{ sample} & & \text{ordered } U(0,1) \text{ sample} \end{array}$$

it can be shown from (2.1) that

$$U_{(i)} = 1 - \prod_{j=1}^i (1 - U_j')^{1/(n-j+1)}, \quad i = 1, \dots, n \quad (2.3)$$

and from (3.1),

$$\begin{array}{ccc} U_1'', U_2'', \dots, U_n'' & \xrightarrow{\quad} & \boxed{R_B^{-1}} \xrightarrow{\quad} U_{(1)}, U_{(2)}, \dots, U_{(n)} \\ \text{Unordered } U(0,1) \text{ sample} & & \text{ordered } U(0,1) \text{ sample} \end{array}$$

where

$$U_{(i)} = \prod_{j=i}^n (U_j'')^{1/j}, \quad i = 1, \dots, n. \quad (2.4)$$

These transforms R_F^{-1} and R_B^{-1} are defined because they "recover" the original ordered sample if this was transformed by R_F or R_B respectively.

3. APPLICATION TO GOODNESS OF FIT WITH TYPE II CENSORED DATA.

3.1. Transforms T_F and T_B : censored samples to full samples.

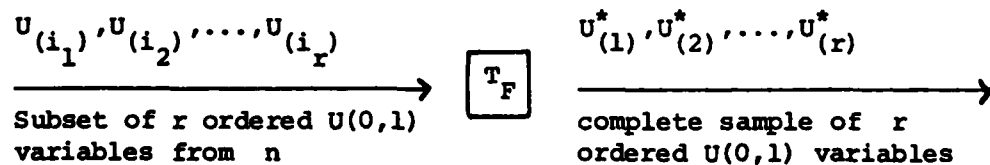
Suppose a censored sample of X -values is given,

$X_{(i_1)} < X_{(i_2)}, \dots, X_{(i_r)}$, and we want to test that they came from a

specified continuous distribution $F(x)$. First, the corresponding censored sample of uniforms is obtained by $U_{(i_1)} = F(X_{(i_1)})$, $U_{(i_2)} = F(X_{(i_2)})$, etc.

The U variables have a joint absolutely continuous distribution so that Rosenblatt's transformation can be applied to them providing a complete unordered uniform sample U'_1, U'_2, \dots, U'_r .

Suppose Rosenblatt's transformation is applied forward, namely first mapping $U_{(i_1)}$ with its marginal distribution, then mapping $U_{(i_2)}$ with its conditional distribution given $U_{(i_1)}$, and so on, to give the set U' ; then apply R_F^{-1} to the U' variables and obtain a set $U^*_{(1)}, U^*_{(2)}, \dots, U^*_{(r)}$ which is a complete ordered $U(0,1)$ sample. Symbolically let this resulting transformation be T_F :



Thus T_F transforms any ordered subset of uniforms into a new complete ordered set of uniforms.

In a similar way, one can define T_B , by first applying Rosenblatt's transformation to $U_{(i_1)}, U_{(i_2)}, \dots, U_{(i_r)}$ but starting with

$U_{(i_r)}$, continuing with $U_{(i_{r-1})}$ given $U_{(i_r)}$, etc., and then applying R_B^{-1} , to give the final ordered set U^* .

The characterizing properties of the distribution of the order statistics make both T_F and T_B doubly invariant characterizations in the sense that they map order statistics into order statistics and the set $U_{(i_1)}, \dots, U_{(i_r)}$ has the joint distribution of the corresponding subset of ordered uniforms if and only if $U_{(1)}^*, \dots, U_{(r)}^*$ is distributed as an ordered complete sample of uniforms. In O'Reilly and Stephens (1982), in connection with transformations of exponentials to uniforms, these concepts were explored. There it was found that doubly invariant characterizations were the basis of procedures with good power properties. It is in this spirit that we now suggest T_F or T_B for censored data.

Example. In order to apply either T_F or T_B in practice one requires knowledge of marginals and conditionals of uniform order statistics. These have nice properties and so the application is quite simple.

Suppose for example that one has only $U_{(2)}$, $U_{(5)}$ and $U_{(7)}$ out of a sample of size 10. In order to apply T_F , one first needs the marginal distribution of $U_{(2)}$, which is a Beta (2, 9) (denoted $B_{2,9}(\cdot)$).

$$\text{So } U_1' = B_{2,9}(U_{(2)}) ;$$

then one needs the conditional distribution of $U_{(5)}$ given $U_{(2)}$. Given $U_{(2)}$, the variables $U_{(3)}, U_{(4)}, \dots, U_{(10)}$ behave like an ordered uniform sample of size 8 on the interval $(U_{(2)}, 1)$, so $U_{(5)}$ given $U_{(2)}$ has the same distribution as the third order statistic of a sample of size 8 on the interval $(U_{(2)}, 1)$.

$$\text{So } U_2' = B_{3,6}\{(U_{(5)} - U_{(2)})/(1 - U_{(2)})\} .$$

Finally we need the conditional distribution of $U_{(7)}$ given $U_{(2)}$ and $U_{(5)}$. This is the same as the conditional distribution of $U_{(7)}$ given $U_{(5)}$ because of the Markovian property of order statistics.

Given $U_{(5)}$, the order statistics $U_{(6)}, U_{(7)}, \dots, U_{(10)}$ have the same distribution as a uniform ordered sample of size 5 on the interval $(U_{(5)}, 1)$.

So $U'_3 = B_{2,4} \{ (U_{(7)} - U_{(5)}) / (1 - U_{(5)}) \}$.

With U'_1, U'_2, U'_3 so computed one would finally apply R_F^{-1} which yields:

$$U_{(1)}^* = 1 - (1 - U'_1)^{1/3}$$

$$U_{(2)}^* = 1 - (1 - U'_1)^{1/3} (1 - U'_2)^{1/2}$$

$$U_{(3)}^* = 1 - (1 - U'_1)^{1/3} (1 - U'_2)^{1/2} (1 - U'_3)^{1/1}.$$

For the right-censored sample, and also for the doubly-censored sample, which often occur in practice, T_F and T_B are straightforward, as follows.

3.2. Right-Censored Case.

For $U_{(1)}, U_{(2)}, \dots, U_{(r)}$ out of n , the application of Rosenblatt's transformation in a forward ordering yields:

U'_1, U'_2, \dots, U'_r independent $U(0,1)$ variables with

$$U'_i = 1 - \{ (1 - U_{(i)}) / (1 - U_{(i-1)}) \}^{n-i+1}, \quad i = 1, \dots, r, \quad U_{(0)} = 0;$$

then application of R_F^{-1} (Equation 2.3) to U_1^i, \dots, U_r^i yields the desired complete uniform sample of size r , $U_{(1)}^*, \dots, U_{(r)}^*$.

If instead we use T_B , first we apply Rosenblatt's transformation in a backward ordering which yields:

$$U_1'', \dots, U_r'' \text{ iid } U(0,1), \text{ where}$$

$$U_i'' = (U_{(i)} / U_{(i+1)})^i, \quad i = 1, \dots, r-1; \text{ and}$$

$$U_r'' = B_{r,n-r+1}(U_{(r)}).$$

Finally R_B^{-1} (Equation 2.4) is applied to U_1'', \dots, U_r'' which results in a complete ordered uniform sample of size r .

For this case T_F does not require an evaluation of a Beta distribution function; this gives a computational advantage over T_B .

Note that for the right-censoring case, T_B reconstructs the random variables given in Theorem 1 of Michael and Schucany (1979).

In a similar way, a left censored sample can be transformed; in this case T_F reconstructs the procedure of Michael and Schucany.

3.3. Doubly-censored case.

Suppose the available order statistics from the uniform sample of size n are:

$$U_{(k+1)}, U_{(k+2)}, \dots, U_{(k+r)}$$

where $1 < k < n-r$.

For T_F , first Rosenblatt's transformation is applied starting with the marginal of $U_{(k+1)}$, then with the conditional of $U_{(k+2)}$ given $U_{(k+1)}$, and so on, which yields

$$U'_1 = B_{k+1, n-k}(U_{(k+1)}) \text{ and}$$

$$U'_i = 1 - \{(1 - U_{(k+i)}) / (1 - U_{(k+i-1)})\}^{n-k-i+1}, \quad i = 2, \dots, r$$

then R_F^{-1} is applied to U'_1, \dots, U'_r .

For T_B , Rosenblatt's transformation yields

$$U''_r = B_{k+r, n-k-r+1}(U_{(k+r)})$$

$$U''_i = \{U_{(k+i)} / U_{(k+i+1)}\}^{k+i}, \quad i = 1, 2, \dots, r-1$$

then R_B^{-1} is applied to U''_1, \dots, U''_r .

In this censoring scheme there seems to be no computational advantage of T_F over T_B since both require the Beta distribution.

3.4. Numerical examples.

The following examples were taken from Figure 1, p. 437 of Michael and Schucany. The values of the actual observations are interpreted from the plots.

In the three examples that follow $n = 9$ and $r = 5$, and the censoring is right-censoring.

(a) For the first example $U_{(1)}, \dots, U_{(5)}$ are 0.1, 0.2, 0.3, 0.4 and 0.5. T_F , as described in Section 3.2, gives a complete ordered uniform set U^* with values 0.1211, 0.3014, 0.4785, 0.6536 and 0.8272; the Anderson-Darling statistic is $A^2 = 0.1884$. The formula for A^2 , for r ordered values U^* is

$$A^2 = -r - \frac{1}{r} \left[\sum_{i=1}^r (2i-1) \{ \log_e (U_{(r)}^*) + \log_e (1 - U_{(r+1-i)}^*) \} \right].$$

Transformation T_B gives another complete uniform set 0.1741, 0.3482, 0.5223, 0.6964 and 0.8706 with $A^2 = 0.1947$.

(b) For the second example, $U_{(1)}, \dots, U_{(5)}$ are 0.0206, 0.0412, 0.0618, 0.0824 and 0.1030. T_F gives the set U^* : 0.7329, 0.8210, 0.8775, 0.9231 and 0.9632, with $A^2 = 4.5746$ and T_B gives U^* : 0.0504, 0.1009, 0.1514, 0.2018 and 0.2523, with $A^2 = 4.1808$.

(c) In the third example, $U_{(1)}, \dots, U_{(5)}$ are 0.1794, 0.3588, 0.5382, 0.7176 and 0.8970. T_F gives U^* : 0.0003, 0.0455, 0.7989, 0.4277 and 0.7005 with $A^2 = 2.789$, and T_B gives U^* : 0.1999, 0.3999, 0.5998, 0.7998 and 0.9998, with $A^2 = 1.3926$.

In example (c), there is a noticeable difference in applying T_F or T_B ; the p-value for A^2 after T_F is below 0.05 and the p-value for A^2 after T_B is above 0.15.

4. POWER STUDY.

Four different methods for testing uniformity from a type II-censored sample were considered. These are T_F , T_B and Michael and Schucany's transformation (MS) each followed by A^2 and also Pettitt and Stephens' (1976) modified A^2 (called PS). This last method calculates A^2 directly from the censored sample, without any transformation, and must be compared to a table appropriate to the censoring being considered. Tables are given in Pettitt and Stephens (1976). Other statistics were shown to be, on the whole, inferior to MS and PS by Michael and Schucany (1979); see their comment on p. 438.

The alternative distributions considered in the power study have been used in previous studies for uniformity by Stephens (1974), Quesenberry and Miller (1977), Michael and Schucany (1979) and Dudewicz and Van der Muelen (1981).

These alternative non-uniforms are the following:

- F_1 : the distribution of Z^2 where $Z \sim U(0,1)$.
- F_2 : the distribution of $1 - Z^2$, where $Z \sim U(0,1)$.
- F_3 : the distribution of $0.5Z_1 + 0.5Z_2$, where Z_1, Z_2 are independently $U(0,1)$.
- F_4 : a mixture of F_1 and F_2 with equal weights.
- F_5 : the distribution of $Z - 0.5$ (if $Z > 0.5$) or $Z + 0.5$ (if $Z < 0.5$)
where $Z \sim F_3$.

The different censoring schemes studied were condensed in two separate tables:

(a) The first censoring scheme is of a sample of size 20 censored at both extremes, thus only

$$U_{(k+1)}, U_{(k+2)}, \dots, U_{(k+r)}$$

were observed.

The number r of available observations was taken to be 5, 10 or 15, and k , which determines the asymmetry of the censoring was varied such that $p = k/(20-r)$ was 0.0, 0.2, 0.4, 0.6, 0.8 and 1.0. Thus p represents the fraction of censored observations at the left; $p = 0$ means right censoring whereas $p = .2$ means that 20% of the censored values occurred at the left. The power results for this censoring scheme and the different alternatives appear in table 1.

(b) In the second censoring scheme, the sample is censored in the middle, thus only the r values

$$U_{(1)}, U_{(2)}, \dots, U_{(\ell)} \quad \text{and} \quad U_{(20-r+\ell+1)}, \dots, U_{(20)}$$

were observed.

The number r of available observations is again varied as 5, 10 and 15. The number ℓ , of available observations at the left, was varied such that $q = \ell/r$ was 0.0, 0.2, 0.4, 0.6, 0.8 and 1.0. The power results appear in table 2.

For both tables, 1000 samples were generated and the percentage of times that the corresponding test detected the alternative, was recorded. The size of the tests was in all cases $\alpha = 0.05$.

For the first censoring scheme, that of censoring at both extremes, except for the cases where p is 0.0 or 1.0 and which correspond to right and left-censoring respectively, the MS transformation could be applied in two different ways (see their comments on p. 439, first paragraph). Both possibilities were considered in table 1 with the results recorded in the same cell. It can be observed that both possibilities lead to similar power.

Power studies for samples censored at both extremes.

p = fraction censored at left = $k/(20-r)$. Size of test $\alpha = 0.05$.

Alternative $\frac{P}{r}$	T_F Followed by A^2								T_B Followed by A^2								PS								
	0	.2	14	16	.8	1	0	.2	.4	.6	.8	.1	0	.2	.4	.6	.8	.1	0	.2	.4	.6	.8	1	
F_1	5	81	74/70	59/52	48/41	35/30	22	81	66	52	41	29	22	81	78	74	65	55	36	39	69	65	61	59	48
	10	83	70/68	63/59	51/47	40/37	27	85	71	60	47	36	27	83	80	75	71	66	58	71	76	74	73	69	67
	15	84	76/76	73/72	61/60	52/51	49	84	77	70	63	57	49	84	82	80	77	75	73	79	86	87	86	85	79
F_2	5	22	32/37	45/50	55/60	68/73	81	36	55	65	74	78	81	22	29	41	52	66	81	48	59	61	65	69	39
	10	27	39/41	45/49	55/60	72/74	83	58	66	71	75	80	83	27	36	47	60	71	85	67	69	73	74	76	71
	15	49	55/56	65/66	70/70	78/78	84	73	75	77	80	82	84	49	57	63	70	77	84	79	85	86	87	86	79
F_3	5	31	16/13	4/1	1/3	13/17	31	25	2	0	6	19	31	31	19	6	0	2	25	38	11	0	0	11	38
	10	30	16/12	3/2	2/4	11/14	30	12	2	1	8	20	30	30	20	8	1	2	12	13	0	0	0	0	13
	15	21	9/8	5/4	5/5	9/10	21	8	3	4	8	14	21	21	14	8	4	3	8	4	4	0	0	4	4
F_4	5	36	21/20	12/12	14/15	21/23	36	38	20	15	19	26	36	36	26	19	15	20	38	17	30	22	22	30	17
	10	34	19/19	14/14	15/15	21/24	34	30	20	18	22	27	34	34	27	22	18	20	30	27	27	26	26	27	27
	15	30	23/23	22/22	19/19	22/22	30	32	26	24	25	27	30	30	27	25	24	26	32	31	47	48	48	47	31
F_5	5	15	28/27	43/41	40/43	29/31	15	19	28	45	50	29	15	15	29	50	45	28	19	3	21	45	45	21	3
	10	35	48/45	47/47	49/50	46/48	35	35	41	47	50	48	35	35	48	50	47	41	35	25	38	48	48	38	25
	15	45	43/42	46/46	42/42	44/44	45	44	44	43	43	43	45	45	43	43	43	44	44	43	61	62	62	61	43
Test procedure																									

TABLE 2

Power study for samples censored in the middle.

Sample size $n = 20$. Observations available are $X_{(1)}, \dots, X_{(\ell)}$ and $X_{(20-r+\ell+1)}, \dots, X_{(20)}$; r = number of available observations; $q = \ell/r$ = fraction available at left end. Size of test $\alpha = 0.05$.

Test Procedure	Alternative $r \backslash q$	MS Followed by A^2										T_F Followed by A^2										T_B Followed by A^2										PS									
		0	.2	.4	.6	.8	1	0	.2	.4	.6	.8	1	0	.2	.4	.6	.8	1	0	.2	.4	.6	.8	1	0	.2	.4	.6	.8	1										
F_1	5	22	50	69	76	79	81	22	51	71	71	82	81	36	44	59	70	75	81	48	38	32	29	29	39	48	38	32	29	29	39										
	10	27	67	79	83	84	83	27	66	79	84	84	85	58	66	74	78	81	83	67	59	59	66	68	71	67	59	59	66	68	71										
	15	49	77	82	83	83	84	49	77	82	83	84	84	73	77	80	81	82	84	79	72	75	79	78	79	79	72	75	79	78	79										
F_2	5	81	79	76	69	50	22	81	75	70	59	44	36	81	82	77	71	51	22	39	29	29	32	38	48	39	29	29	32	38	48										
	10	83	84	83	79	67	27	83	81	78	74	66	58	85	84	84	79	66	27	71	68	66	59	59	67	71	68	66	59	59	67										
	15	84	83	83	82	77	49	84	82	81	80	77	73	84	84	83	82	77	49	79	78	79	75	72	78	79	78	79	75	72	78										
F_3	5	31	4	2	2	4	31	31	10	16	21	25	25	25	25	21	16	10	31	38	53	66	66	53	38	38	53	66	66	53	38										
	10	30	4	3	3	4	30	30	8	10	10	12	12	12	12	10	10	8	30	13	37	54	54	37	13	13	37	54	54	37	13										
	15	21	8	5	5	8	21	21	10	8	7	7	8	8	7	7	8	10	21	4	14	30	30	14	4	4	14	30	30	14	4										
F_4	5	36	35	37	37	35	36	36	32	36	39	38	38	38	38	39	36	32	36	17	9	5	5	9	17	17	9	5	5	9	17										
	10	34	35	40	35	40	34	34	34	40	40	36	30	30	36	40	40	34	34	27	22	25	25	22	27	27	22	25	25	22	27										
	15	30	36	40	40	36	30	30	36	41	38	35	32	32	35	38	41	36	30	31	26	35	35	26	31	31	26	35	35	26	31										
F_5	5	15	8	4	4	8	15	15	6	5	6	11	19	19	11	6	5	6	15	27	1	0	0	1	27	27	1	0	0	1	27										
	10	35	16	9	9	16	35	35	12	6	14	27	35	35	27	14	6	12	35	25	11	3	3	11	25	25	11	3	3	11	25										
	15	45	33	20	20	33	45	45	31	16	28	41	44	44	41	28	16	31	45	43	28	16	16	28	43	43	28	16	16	28	43										

Comments on Table 1.

When $p = 0$ (the sample is right censored), the results agree with those reported by Michael and Schucany (1979), to within Monte Carlo variation. For this case, ($p = 0$), the Pettitt-Stephens (1976) procedure dominates under alternative F_2 where the second best is T_F (recall that in this case T_B is equivalent to MS and also T_F has a computational advantage over T_B). Against F_3 , T_B dominates T_F and except for highly censored samples ($r = 5$), also dominates PS.

For the rest of the alternatives, still with $p = 0$, T_B and T_F are roughly equivalent and dominate PS.

Under left censoring, that is when $p = 1$, procedure PS outperforms the rest under alternative F_1 where T_B is the second best (here, MS and T_F coincide). Under F_3 , except for highly censored samples ($r = 5$), where PS does well, T_F dominates T_B and PS.

For the rest of the alternatives, still with $p = 1$, T_B and T_F are again roughly equivalent and dominate PS.

If censoring occurs at both extremes, then the relative performances of the test procedure depend on the degree of symmetry of the censorship, that is, how close p is to 0.5.

Against F_4 , procedure PS does very well except for values of p on the extremes. Against F_3 however, PS is outperformed by MS, which in turn is outperformed by T_F when there is more censoring at the right and by T_B if there is more censoring at the left.

Against F_1 , PS and T_B behave similarly and do better than MS or T_F . Against F_2 , PS and T_F are similar and outperform the other two.

Against F_5 , there are no big discrepancies, except if $r = 15$ (25% censoring) in which case PS does better.

Comments on Table 2.

Consider the case where $q = .2, .4, .6$ or $.8$ to avoid the cases of left- and right-censoring ($q = 0$ and $q = 1$), already discussed.

Against F_1 , MS and T_F dominate. Against F_2 , MS and T_B dominate and against F_3 , PS dominates. Against F_4 , MS, T_F and T_B are comparable and PS behaves poorly, especially with high censorship (75%).

Against F_5 , T_F or T_B do well if the largest amount of available information is at the left or right respectively. The procedure MS behaves symmetrically in this respect and PS yields low power if censorship is above 60%.

5. APPLICATION TO GOODNESS OF FIT WITH TYPE 1 CENSORED DATA.

In Section 3, the transformations T_F and T_B were defined as those resulting from Rosenblatt's transformation applied to the available subset of the order statistics in a forward or backward ordering, followed by R_F^{-1} and R_B^{-1} respectively.

For one sided or double censoring of type I, one can still find T_B and T_F . For this we need only to consider a conditionality argument.

5.1. Right censored case (Type I).

The observations $U_{(1)}, U_{(2)}, \dots, U_{(r)}$ are the sample values that were less than a fixed and known censoring constant t . Observe that the integer r is random.

Define the event C to be $U_{(r)} < t < U_{(r+1)}$.

Consider the conditional joint distribution of $U_{(1)}, \dots, U_{(r)}$ given C . Since this is absolutely continuous (almost surely), one can apply Rosenblatt's transformation in a forward or backward ordering. For example, the transformation T_B works as follows.

First, the conditional distribution of $U_{(r)}$ given C is needed to map $U_{(r)}$, then the conditional distribution of $U_{(r-1)}$ given C and $U_{(r)}$ is needed to map $U_{(r-1)}$, then the conditional distribution of $U_{(r-2)}$ is needed to map $U_{(r-1)}$, then the conditional distribution of $U_{(r-2)}$ given C , $U_{(r)}$ and $U_{(r-1)}$ is required to map $U_{(r-2)}$, and so on.

The conditional distribution $S(\cdot)$ of $U_{(r)}$ given C is

$$S_{U_{(r)}}(u|C) = \begin{cases} 0 & \text{if } u \leq 0 \\ \left(\frac{u}{t}\right)^r & \text{if } u \in (0, t) \\ 1 & \text{if } u \geq t. \end{cases}$$

Similarly, the conditional distribution of $U_{(r-1)}$ given C and $U_{(r)}$ is found to be

$$S_{U_{(r-1)}}(u|C, U_{(r)}) = \begin{cases} 0 & \text{if } u \leq 0 \\ \left(\frac{u}{U_{(r)}}\right)^{r-1} & \text{if } u \in (0, U_{(r)}) \\ 1 & \text{if } u > U_{(r)} \end{cases}$$

Finally, T_B consists in applying R_B^{-1} (formula 2.4) to the set

U_1'', \dots, U_r'' given by

$$U_r'' = \{U_{(r)}/t\}^r$$

and

(5.1)

$$U_j'' = (U_{(j)}/U_{(j+1)})^j, \quad j = 1, \dots, r-1$$

The resulting U'' are, conditionally on C , independent $U(0,1)$, hence after applying R_B^{-1} to these, we obtain $U_{(1)}^*, \dots, U_{(r)}^*$ which, conditionally on C , are distributed like a complete ordered $U(0,1)$ sample.

A similar derivation yields T_F in this Type I right censored case. The Rosenblatt transformation gives U_1', \dots, U_r' where

$$U_1' = 1 - \left\{ 1 - \frac{U_{(1)}}{t} \right\}^r$$

and (5.2)

$$U_i' = 1 - \left\{ 1 - (U_{(i)} - U_{(i-1)}) / (t - U_{(i-1)}) \right\}^{r-i+1}, \quad i = 2, \dots, r;$$

then R_F^{-1} is applied to the U_i' to give the ordered set U^* .

5.2. Double Censored case (Type I).

The uncensored observations $U_{(k+1)}, U_{(k+2)}, \dots, U_{(k+r)}$ are the sample values which were less than a fixed known censoring constant t_2 and greater than another fixed and known censoring constant t_1 , with $t_1 < t_2$.

Applying Rosenblatt's transformation conditional on the events C and D defined by

$$C = [U_{(k+r)} < t_2 < U_{(k+r+1)}], \quad D = [U_{(k)} < t_1 < U_{(k+1)}],$$

the following results are obtained.

For T_B , the U_1'', \dots, U_r'' needed before R_B^{-1} is applied are given by

$$U_r'' = \{(U_{(k+r)} - t_1)/(t_2 - t_1)\}^x \quad (5.3)$$

and

$$U_i'' = \{(U_{(k+i)} - t_1)/(U_{(k+i+1)} - t_1)\}^i, \quad i = 1, \dots, r-1.$$

For T_F , the U_1', \dots, U_r' needed before applying R_F^{-1} are given by

$$U_r' = 1 - \{(t_2 - U_{(k+1)})/(t_2 - t_1)\}^x$$

(5.4)

and

$$U_i' = 1 - \{(t_2 - U_{(k+r-i+1)})/(t_2 - U_{(k+r-i)})\}^i$$

for $i = 1, \dots, r-1$.

In this type of censoring situation (and similarly for Type 1 right- or left-censoring) one finishes with a set of random variables $U_{(1)}^*, \dots, U_{(r)}^*$ which conditionally on C and D has the same distribution as a complete ordered $U(0,1)$ sample of size r . The Anderson-Darling A^2 computed from these random variables, conditionally on C and D , has a well known distribution, which in the upper tail is, for all practical purposes, independent of r (if $r \geq 5$; see Stephens, 1974); hence the test statistic A^2 applied after T_B or T_F in a type I censoring situation is essentially distributed as in the usual unconditional test for uniformity.

6. GENERAL COMMENTS.

(a) For Type 2 censoring, there seems to be no overall best procedure from the comparison carried out in Section 4. It is true, however, that in a practical situation one knows the type of censoring which is confronted and if one suspects the alternative, then the general comparisons given should be useful.

(b) An appealing property of the procedures T_F and T_B is that they indicate explicitly how a given subset of an ordered uniform sample may be transformed to a complete ordered uniform sample, by a method which applies in great generality.

(c) When defining T_F in Section 3, one first maps the available order statistics with Rosenblatt's transformation starting with the smallest, then the next to the smallest and so on. At this stage one has an independent $U(0,1)$ sample. Any fixed permutation of this independent sample could then be transformed with R_F^{-1} . Theoretically this is a valid procedure but in practice, we found that the power was highest when the ordering was maintained. This empirical result suggested the use of R_F^{-1} when the Rosenblatt transformation was used forward and R_B^{-1} when it was used with the backward ordering. In this way one maintains the identity of the observations. This agrees with an observation concerning the retention of identity that appears in Michael and Schucany (1979, p. 439, in the first paragraph).

(d) For Type 1 censoring, there is extra information in the random number r (and, for double censoring, in k). For example, with only right censoring, a value r far from its expected value nt indicates lack

of uniformity, even if the r values themselves were evenly distributed in the interval $(0,t)$. The need to make a test involving r has been discussed by Stephens, and by Michael and Schucany, in Chapters 5 and 12 of D'Agostino and Stephens (1985). In Chapter 5 it is suggested that a test that r is reasonable (r has a binomial distribution with parameters n and t) be combined with the test for uniformity. A two-stage procedure of this type has recently been discussed by Maag and Dufour (1985). An interesting question which then arises is how best to choose the α -levels of the component tests to give the desired overall α -level. In Chapter 12 of D'Agostino and Stephens (1985), Michael and Schucany point out that the Pettitt-Stephens procedure makes use of r ; so also does a statistic based on the spacings between the $U_{(i)}$.

Further work is needed to compare these several methods of incorporating the information contained in r , and, for double censoring, in k .

ACKNOWLEDGEMENTS

This work was supported by the Natural Science and Engineering Research Council of Canada and was done while the first author was a visitor at the Department of Mathematics and Statistics at Simon Fraser University.

REFERENCES

- [1] D'Agostino, R.B. and Stephens, M.A. (1985). Goodness-of-Fit Techniques. New York: Marcel Dekker.
- [2] Dudewicz, R.J. and Van der Meulan, E.C., (1981). Entropy based tests of uniformity. J. Amer. Statist. Assoc., 76, 967-974.
- [3] Maag, V. and Dufour, R. (1985). Two stage goodness-of-fit for truncated or censored samples. Contributed paper, Annual Meeting, Statistical Society of Canada, 1985.
- [4] Malmquist, S., (1950). On a property of order statistics from a rectangular distribution. Skand. Aktuariidskrift, 33, 214-222.
- [5] Michael, J.R. and Schucany, W.R., (1979). A new approach to testing goodness of fit for censored samples. Technometrics, 21, 435-441.
- [6] O'Reilly, F.J. and Stephens, M.A., (1982). Characterizations and goodness of fit tests. J.R. Statist. Soc., B, 44, 353-369.
- [7] Pettitt, A.N. and Stephens, M.A., (1976). Modified Cramer von Mises statistics for censored data. Biometrika, 63, 291-298.
- [8] Quesenberry, C.P. and Miller, F.L., Jr., (1977). Power studies of some tests for uniformity. J. Statist. Comput. Simulation, 5, 169-191.
- [9] Rosenblatt, M., (1952). Remarks on a multivariate transformation. Ann. Math. Statist. 28, 470-472.
- [10] Stephens, M.A., (1974). EDF Statistics for goodness-of-fit and some comparisons. J. Amer. Statist. Assoc., 69, 730-737.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM	
1. REPORT NUMBER 361	2. GOVT ACCESSION NO. 411-4157 957	3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) Transforming Censored Samples And Testing Fit		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT	
		6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s) F. J. O'Reilly and M. A. Stephens		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0475	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267	
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 411SP		12. REPORT DATE July 23, 1985	
		13. NUMBER OF PAGES 28	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED	
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED			
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)			
18. SUPPLEMENTARY NOTES			
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Goodness-of-fit tests, Uniform distributions, Censored samples.			
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) One approach to testing the goodness-of-fit of a completely specified continuous distribution with a censored sample, is to transform the corresponding uniform censored sample into a complete uniform sample and then to use any of the classical uniformity tests. Two systematic procedures for transforming a censored uniform sample are proposed. A Monte Carlo power study was conducted to analyze the relative merits of these procedures when followed by the Anderson-Darling A^2 test for uniformity. The suggested transformations are versatile and with A^2 , give good power for most alternatives.			

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

9-85

DTIC